Deforestation Monitoring in the Colombian Amazon: Semantic Segmentation and Change Detection with Deep Learning and Sentinel-2.

A extended research project report submitted to the University of Manchester for the degree of MSc. Data Science in the Faculty of Humanities.

Submission date: September 2025

Student ID: 11513507

School of Social Sciences

TABLE OF CONTENTS

Word count: 6006

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTELLECTUAL PROPERTY STATEMENT

i. The author of this extended research project report (including any appendices and/or schedules to this report) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes. ii. iii. iv.

ii. Copies of this report, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered into. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks, and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the report, for example graphs and tables ("Reproductions"), which may be described in this report, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this report, the Copyright and any Intellectual Property and/or Reproductions described in it may 17 MSc Data Science take place is available in the University IP Policy (see https://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant dissertation restriction declarations deposited in the University Library, The University Library's regulations (see https://www.library.manchester.ac.uk/about/regulations/) and in The University's Guidance for the Presentation of dissertations.

DECLARATION

No portion of the work referred to in this extended research project report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

ABSTRACT

Tropical forests are under severe threat from deforestation, and traditional monitoring methods are inadequate for the scale and urgency of this crisis. The application of deep learning for this task is often hindered by a critical bottleneck: the scarcity of large-scale, accurately labelled training data. This study directly confronts this challenge by developing and implementing a novel, automated pipeline that leverages Google Earth Engine to process Sentinel-2 satellite imagery and Google's Dynamic World dataset to generate pseudo ground-truth labels. To address severe class imbalance that caused initial model failure, a nuanced 4-class land cover problem (Trees, Grass, Shrub/Scrub, Other) was formulated with inverse frequency class weighting, before comparing a standard U-Net and an Attention U-Net architecture. The results demonstrate that the Attention U-Net outperformed the standard U-Net, achieving a peak validation Mean Intersection-over-Union (IoU) of 0.674 compared to 0.609 for the standard U-Net. In the final binary (Forest vs. Non-Forest) evaluation, the Attention U-Net delivered more balanced and effective performance, with a Forest F1-score of 0.93 and a recall of 0.90, compared to the standard U-Net's F1-score of 0.91 and recall of 0.85. The validated Attention U-Net was then applied to a 4,430 km² deforestation hotspot in the San José del Guaviare region, quantifying a net forest loss of 177.30 km² between 2021 and 2023. This work establishes a complete, scalable, and reproducible workflow that can be adapted for near-real-time deforestation monitoring, providing a foundational tool to support timely conservation interventions and strategic policymaking.

1. INTRODUCTION

Tropical forests, often described as the "lungs of the planet," are critical reservoirs of biodiversity and essential regulators of the global climate (UNEP and FAO, 2020). Among the nations gifted with this vital resource, Colombia stands out as one of the most biodiverse countries per square kilometre, hosting nearly 10% of the world's species within its varied ecosystems, where nearly 1,900 species of birds, 67 mammals, and more than 130,000 plant species are found (Procolombia, 2023). The Colombian Amazon, with a size of 483,119 km^2 and covering 42% of the national territory, is a cornerstone of this natural heritage, containing an extraordinary wealth of life, and serving as the ancestral home for numerous Indigenous communities (SIAT, 2023).

Despite their ecological and cultural significance, these forests are under severe and escalating threat. According to data from Colombia's Institute of Hydrology, Meteorology and Environmental Studies (IDEAM, 2024) the nation has been grappling with alarming deforestation rates, with 123,517 hectares lost in 2022 alone, an area roughly equivalent to the size of Los Angeles or 1,200 km^2. This clearing, driven by the expansion of agriculture, cattle ranching, illegal mining, and illicit crops, not only decimates biodiversity but also jeopardises the stability of the Amazonian biome and its role in hemispheric climate systems (Molina, 2024). Effective and timely monitoring of these vast and often inaccessible regions is therefore not just a national priority but a global imperative.

The conventional approach to forest monitoring, reliant on map interpretation, manual ground surveys and ancillary data analysis, is incapable of addressing the scale and urgency of this crisis. Such methods are labour-intensive, and prohibitively expensive (Alshehri et al. 2024). In response, the scientific and conservation communities have turned to remote sensing, using satellite imagery to observe and quantify land cover change across broad regions (James, 2025). The paradigm shifted in 2008 when the entire Landsat archive was made available to the public at no cost, a move that unleashed the potential of the data for science and monitoring (Wulder et al., 2016). This was complemented by the Copernicus programme's Sentinel satellites, which were designed specifically to provide systematic, operational data for services like environmental protection (Drusch et al., 2012). The resulting petabyte-scale archives necessitated new cloud-based platforms like Google Earth Engine to make analysis possible (Gorelick et al., 2017). Consequently, this data now underpins global-scale research, from mapping worldwide deforestation (Hansen et al., 2013) to tracking long-term changes in

surface water (Pekel et al., 2016), solidifying its role as an essential tool for understanding our planet.

However, the availability of data is only one part of the problem. The primary problem lies in its interpretation. Early remote-sensing techniques often relied on pixel-based analyses of spectral indices like the Normalized Difference Vegetation Index (NDVI) to assess vegetation health (James, 2025). While useful for assessing vegetation health, these pixel-based indices are fundamentally limited. Their spectral calculations are easily distorted by atmospheric effects like haze and aerosols, which can make a healthy forest appear less vigorous (Rech, 2023). Furthermore, as they analyse each pixel in isolation, they cannot leverage spatial context, often leading to misclassifications, such as confusing a patch of dark, moist soil with the shaded canopy of a dense forest.

Subsequent rise of machine learning (ML) models, such as Random Forest (RF) and Support Vector Machines (SVMs), offered an improvement by classifying pixels based on a range of spectral and textural features (Pal, 2005; Mountrakis, Im and Ogole, 2011; Belgiu and Drăguţ, 2016). These models, however, depend heavily on a process of manual "feature engineering," which is not only time-consuming but also requires substantial domain expertise to select the optimal features for a given task.

The limitations of traditional ML have been largely superseded by the paradigm shift brought about by deep learning (DL), particularly the application of Convolutional Neural Networks (CNNs) to image analysis (Lecun et al., 1998; Zhu et al., 2017). Unlike their predecessors, deep learning models, particularly CNNs, have revolutionised the field by automatically learning hierarchical feature representations directly from raw image data, thus eliminating the feature engineering challenge and enabling the detection of intricate spatial patterns frequently achieving unprecedented levels of accuracy. (Zhang et al., 2016; Zhu et al., 2017). Deep learning models excel at semantic segmentation, the task of classifying every pixel in an image to produce a detailed, wall-to-wall map of land cover categories. Architectures like U-Net, known for its encoder-decoder structure (Ronneberger et al., 2015), have become state-of-the-art for segmenting remote-sensing imagery (Lv et al., 2023). Recent research has further pushed the boundaries with Transformer-based (Panboonyuen et al., 2021) and attention mechanisms (Wang et al., 2022), to perform semantic segmentation in remote sensing with high accuracy.

Despite the power of these models, their application in real-world conservation efforts is constrained by a critical bottleneck: the scarcity of large-scale, accurately labelled training data. Supervised deep learning models are data-hungry, requiring thousands of labelled examples to achieve high performance. The process of manually annotating satellite imagery is demanding, time-consuming, and subjective. While regional datasets have been developed, such as the BrazilDAM (Ferreira et al., 2020) dataset for detecting mining tailings dams, they are often tailored to specific geographic areas or land cover types, limiting their transferability to ecologically distinct regions like the Colombian Amazon.

This project directly confronts this data-scarcity challenge by developing and implementing a complete and scalable deep learning pipeline for deforestation analysis in Colombia. The core of this work's innovation lies in its novel data generation strategy, which leverages two powerful, cloud-based platforms. Firstly, it uses Google Earth Engine (GEE) (Gorelick et al., 2017), a planetary-scale platform for geospatial analysis, to access and pre-process years of Sentinel-2 satellite imagery (Drusch et al., 2012), creating cloud-free composites of the study area without the need for prohibitive local computation. Secondly, to overcome the lack of ground truth labels, the pipeline employs Google's Dynamic World dataset (Brown et al., 2022). Dynamic World provides a reliable and consistently updated source of global land-cover classifications that can serve as "pseudo ground truth" labels for training at a large scale.

By pairing the processed Sentinel-2 imagery with labels from Dynamic World, this project automates the creation of a high-quality dataset. This dataset is then used to train and compare two state-of-the-art semantic segmentation models: the classic U-Net and the more recent Attention U-Net architecture. The final output is a high-resolution land-cover map which, when compared across different years, can be used to precisely identify and quantify areas of forest loss. This work therefore delivers an actionable tool for stakeholders to monitor deforestation, particularly in hotspots where access and real-time monitoring is arduous, and to better direct conservation and policy interventions.

The primary goals of this work are threefold:

1. The design and implementation of a novel, automated pipeline combining Google Earth Engine and the Dynamic World dataset to generate large-scale training data for land cover classification, addressing a critical data constraint in deforestation monitoring.

2. A comparative analysis of state-of-the-art deep learning architectures (U-Net and Attention U-Net) for semantic segmentation of land cover in the complex ecosystems of Colombia using Sentinel-2 imagery.

3. The production of a high-resolution land-cover map and a quantitative analysis of forest change for a critical deforestation hotspot, providing actionable insights for local conservation efforts.

This paper is organised as follows: Chapter 2 details the study area, the datasets used for training, the implemented neural network architectures, and the experimental configuration. Chapter 3 presents the results of the model training and the land cover change analysis. Finally, Chapters 4 and 5 provide a discussion of the results and the overall conclusions of the study, respectively.

2. MATERIALS AND METHODS

2.1 STUDY AREA

The geographical focus of this study is the northwestern Colombian Amazon, a region critically important for both its immense biodiversity and its alarming rates of forest loss. This area is located at the "Arc of Deforestation," a frontier where agricultural expansion, infrastructure development, and illicit activities exert intense pressure on rainforest ecosystems. According to Colombia's Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), the Amazon region consistently accounts for the majority of national deforestation, with an estimated 68% of the country's total forest loss concentrated there in recent years.

To effectively train and validate the deep learning models, two distinct Regions of Interest (ROIs) were strategically selected within this critical zone, as depicted in Figure 1.
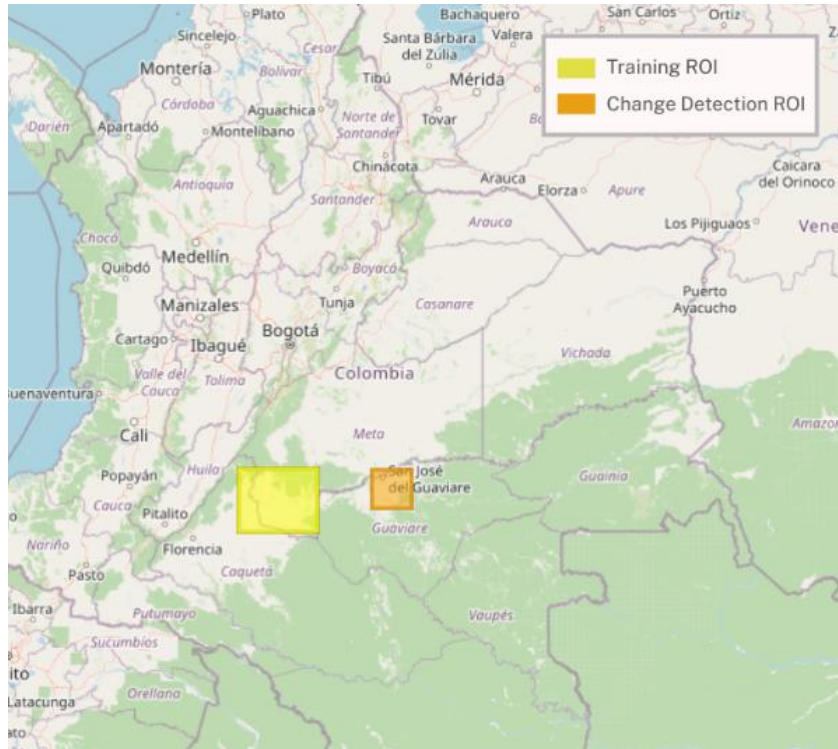
Figure 1. A map showing Colombia with the two ROIs as "Training ROI" in Caquetá/Meta and "Change Detection ROI" in Guaviare.

The primary region for training the semantic segmentation models was a large, diverse area of 15,075.31 km² spanning the municipalities of San Vicente del Caguán (Caquetá department) and La Macarena (Meta department), strategically chosen because it is an active deforestation hotspot containing a diverse mix of landscapes. This includes intact rainforest, established

pastureland, and the critical transitional stages of forest conversion, such as fragmented forest and scrubland.

This diversity is essential for training a robust model, forcing it to learn the spectral signatures of various land cover types and the subtle differences between them. Both Meta and Caquetá are consistently ranked by IDEAM as being among the top five departments for deforestation nationwide, making this ROI a representative sample of the problem being addressed (IDEAM, 2024). The bounding box for this training region is defined by the coordinates: 1.722°N to 2.712°N and 74.889°W to 73.656°W.

For the final change detection analysis, a separate and distinct ROI was selected to test the model's generalization capabilities on a new, unseen area. This 4,430.23 km² region is cantered on the municipality of San José del Guaviare (Guaviare department). Guaviare is another critical deforestation hotspot, often cited as a nexus for land grabbing, road construction, and the expansion of illicit crops that drive forest conversion (UNODC, 2021).

This ROI was chosen as it represents an active and rapidly changing landscape, providing an ideal testbed for quantifying the performance of the trained model in a real-world monitoring scenario. The bounding box for this change detection region is defined by the coordinates: 2.1°N to 2.7°N and 72.8°W to 72.2°W.

2.2 DATASETS

To address the project's objective, two distinct, publicly available satellite data products were used, both sourced via the Google Earth Engine (GEE) cloud platform. The selection of these datasets was based on their spatial and temporal resolution, data quality, and suitability for large-scale land cover analysis and deep learning applications. Figure 2 provides a visual example of how these two datasets align, showing a Sentinel-2 true-color image alongside its corresponding Dynamic World land cover classification mask for a sample area within the study region.

Figure 2. Example of the primary datasets used in the study. (Left) A true-colour composite from Sentinel-2 imagery, which serves as the input features for the model. (Right) The corresponding 10-meter pixel size land cover classification from the Dynamic World.

2.2.1 Sentinel-2 Multispectral Imagery

The primary feature data for this study was sourced from the Sentinel-2 satellite mission of the European Union's Copernicus Earth observation program. We used the Sentinel-2 Level-2A (L2A) surface reflectance collection, which provides analysis-ready, atmospherically corrected imagery. This ensures spectral consistency across time, which is critical for change detection. The mission's high temporal resolution of approximately five days provides a high density of images, enabling the creation of cloud-free composites even in persistently cloudy regions like the Colombian Amazon.

A key advantage of Sentinel-2 is its combination of high spatial resolution and rich spectral detail. The model was trained using a selection of nine bands at 10 and 20-meter pixel size. This included the visible (Red, Green, Blue), Near-Infrared (NIR), and Short-Wave Infrared (SWIR) bands. This spectral range is highly effective for land cover classification (Drusch et al., 2012); the visible bands help identify basic landscape features, while the NIR and SWIR bands are particularly sensitive to vegetation health, structure, and moisture content. This combination provides the model with the comprehensive information needed to accurately differentiate between key classes such as forests, grasslands, water bodies, and urban areas.

2.2.2 Dynamic World Land Cover Data

The ground-truth labels required for model training and validation were derived from the Google Dynamic World V1 dataset. This near-real-time product offers a global, 10-meter pixel size land cover classification, generated by applying a deep learning model to the entire Sentinel-2 satellite archive (Brown et al., 2022). A key advantage of this dataset is its perfect spatial and temporal alignment with the Sentinel-2 feature imagery, eliminating the need for complex data registration or resampling. For every Sentinel-2 scene, Dynamic World provides a per-pixel probability distribution across nine distinct land cover classes: water, trees, grass, flooded vegetation, crops, shrub/scrub, built area, bare ground, and snow/ice.

2.3 NETWORK ARCHITECTURES

To perform the semantic segmentation of land cover, this study implemented and compared two architectures from the U-Net family. The standard U-Net was selected as the foundational model due to its proven effectiveness in producing precise segmentation masks for satellite imagery, where accurate localization of features is essential (Ronneberger et al., 2015). However, a potential limitation of its architecture is the unselective nature of its feature fusion process via skip connections.

To test a direct architectural improvement, the Attention U-Net was chosen as the comparison model (Oktay et al., 2018). This variant introduces attention gates designed to make the feature fusion process more selective and context aware. The primary research question addressed by this comparison is whether this targeted enhancement, focusing on the most relevant spatial features, can yield a significant performance improvement over the baseline U-Net for the complex task of land cover classification

2.3.1 U-Net

The U-Net, originally proposed by Ronneberger et al. (2015), is a fully convolutional neural network designed for end-to-end image segmentation. Its distinctive U-shaped architecture consists of two symmetric paths: a contracting (encoder) path to capture context and an expansive (decoder) path for precise localization. The encoder follows the structure of a typical convolutional network, composed of sequential blocks that each contain two 3x3 convolutional layers with ReLU activations. To mitigate overfitting and improve generalization, Dropout layers are applied within each of these blocks. Each encoder block is followed by a 2x2 max pooling operation for downsampling, which progressively halves the spatial dimensions of the feature maps while doubling the number of feature channels. This process allows the network

to build a hierarchical understanding of the image, with deeper layers learning more complex and abstract features.

The decoder's primary function is to upsample these abstract features and recover the original spatial resolution for pixel-wise classification. A key innovation of the U-Net is the use of skip connections, which concatenate the upsampled feature maps from the decoder with the corresponding high-resolution feature maps from the encoder path. This fusion allows the decoder to leverage both the abstract semantic information from the deep layers and the fine-grained spatial details from the shallow layers, which is crucial for accurately delineating boundaries between land cover classes. The final layer of the network is a 1x1 convolution with a softmax activation function, which produces a probability map for each target class for every pixel. For this study, the U-Net was configured with three downsampling and three upsampling stages, creating a network of four distinct resolution levels.

2.3.2 Attention U-Net

While the standard U-Net is highly effective, its skip connections naively fuse all features from the encoder, potentially propagating redundant or irrelevant information to the decoder. The Attention U-Net, proposed by Oktay et al. (2018), addresses this limitation by integrating attention gates into the skip connections. The attention gate is a sub-network that dynamically generates a weighting mask, or an "attention map," which is applied to the features being passed from the encoder. This map learns to highlight the most salient regions relevant to the segmentation task at a given scale, while suppressing feature responses in irrelevant background areas.

The gate takes two inputs: the high-resolution feature map from the encoder path and a gating signal from the corresponding, deeper decoder path, which provides the necessary contextual information to guide the attention mechanism. By applying this attention map before the concatenation step, the model learns to selectively focus on the most important spatial information, making the feature fusion process more intelligent and efficient. For this project, this allows the model to learn to emphasize the spectral signatures of "Trees" while ignoring nearby, potentially confusing features. The architecture of the Attention U-Net implemented in this study is identical to the baseline U-Net, including the use of Dropout for regularization, with the sole modification being the inclusion of these attention gates at each of the three main skip connection pathways.

2.4 EXPERIMENT SETUP

All model training and experiments were conducted on a high-performance computing system equipped with a 32-core CPU and accelerated by a single NVIDIA Tesla V100-SXM2 GPU with 16 GB of VRAM. The data preparation workflow was executed within the Google Earth Engine (GEE) cloud platform to handle the large-scale geospatial data. For the input features, a cloud-free Sentinel-2 composite was generated for the year 2022 by filtering the `COPERNICUS/S2_SR_HARMONIZED` image collection and applying a per-pixel `median()` reducer. This technique synthesizes the time-series into a single, high-quality image, effectively removing clouds and transient atmospheric noise. Concurrently, a corresponding label image was created from the `GOOGLE/DYNAMICWORLD/V1` collection by applying a `mode()` reducer to find the most frequent land cover classification for each pixel over the same period.

To prepare this data for the deep learning framework, the Sentinel-2 feature bands and the Dynamic World label band were combined into a single multi-band image stack to ensure perfect pixel-wise alignment. The feature bands were then normalized by dividing their surface reflectance values by 3000, scaling them into a range suitable for neural network processing. The final stacked data was exported from GEE as a set of TFRecord files by partitioning the image into 256 x 256 pixel patches. This process yielded a final dataset of 2,279 patches, which was subsequently split into 2,051 for training and 228 for validation.

A key methodological decision was to remap the nine original Dynamic World classes into four target categories relevant to deforestation analysis: Trees, Grass, Shrub/Scrub, and a consolidated Other lass. This strategy was adopted to address the class imbalance that caused initial models to fail. To further mitigate the remaining imbalance, class weights were calculated based on the inverse frequency of each of the four classes within the training set. This strategy penalizes the model more heavily for misclassifying less frequent classes, encouraging a more balanced predictive performance. The models were trained for a maximum of 50 epochs with a batch size of 5 and a shuffle buffer of 250, parameters selected after an iterative experimentation process to balance performance, training stability, and the available computational resources. The Adam optimizer was used for model compilation with its default learning rate, chosen for its adaptive learning capabilities and robust performance across a wide range of deep learning tasks (Kingma and Ba, 2014). To manage memory usage during long runs, validation was performed over 15 steps per epoch. The model's performance was monitored using the 4-class Mean Intersection-over-Union (IoU), with an early stopping patience of 10 epochs to save the best-performing model.

For the final analysis, the model's primary 4-class output was reclassified in a post-processing step to generate a binary Forest vs. Non-Forest map. In this binary scheme, the 'Trees' class was designated as 'Forest', while the 'Grass', 'Shrub/Scrub', and 'Other' classes were aggregated into a 'Non-Forest' category. This two-stage evaluation allows the model to learn from a more nuanced ecological context while providing a final, focused assessment of its efficacy for the ultimate goal of deforestation monitoring. The same data preparation workflow was used to generate two additional inference datasets for the years 2021 and 2023 to serve as the "before" and "after" snapshots for the change detection analysis.

3. RESULTS

This section presents the empirical outcomes of the study, beginning with the training performance of the implemented deep learning models, followed by an evaluation of their predictive accuracy, and culminating in the application of the best-performing model to the change detection analysis.

3.1 MODEL TRAINING AND PERFORMANCE COMPARISON

The U-Net and Attention U-Net models were trained for a maximum of 50 epochs on the 4-class land cover problem. The training process was monitored using the validation Mean Intersection-over-Union (IoU) as the primary performance metric, with an early stopping mechanism to prevent overfitting and select the best-performing model weights.

Figure 3 presents the learning curves for both models, illustrating the validation loss and validation Mean IoU over the course of training. The standard U-Net model trained for 32 epochs before early stopping was triggered, achieving a peak validation Mean IoU of 0.609 at epoch 22. In contrast, the Attention U-Net demonstrated a more stable training progression, training for 40 epochs and reaching a significantly higher peak validation Mean IoU of 0.674 at epoch 30. This suggests that the attention mechanism provided a tangible benefit, enabling the model to learn more effectively from the nuanced 4-class data. The validation loss curves for both architectures show a consistent downward trend without diverging from the training loss, which indicates that both models were successfully optimised without significant overfitting. Despite this, the Attention U-Net consistently maintained a higher IoU score throughout the later stages of training, suggesting that the attention mechanism provided a tangible benefit in learning the nuanced 4-class data and confirming its better performance on the primary multi-class segmentation task.
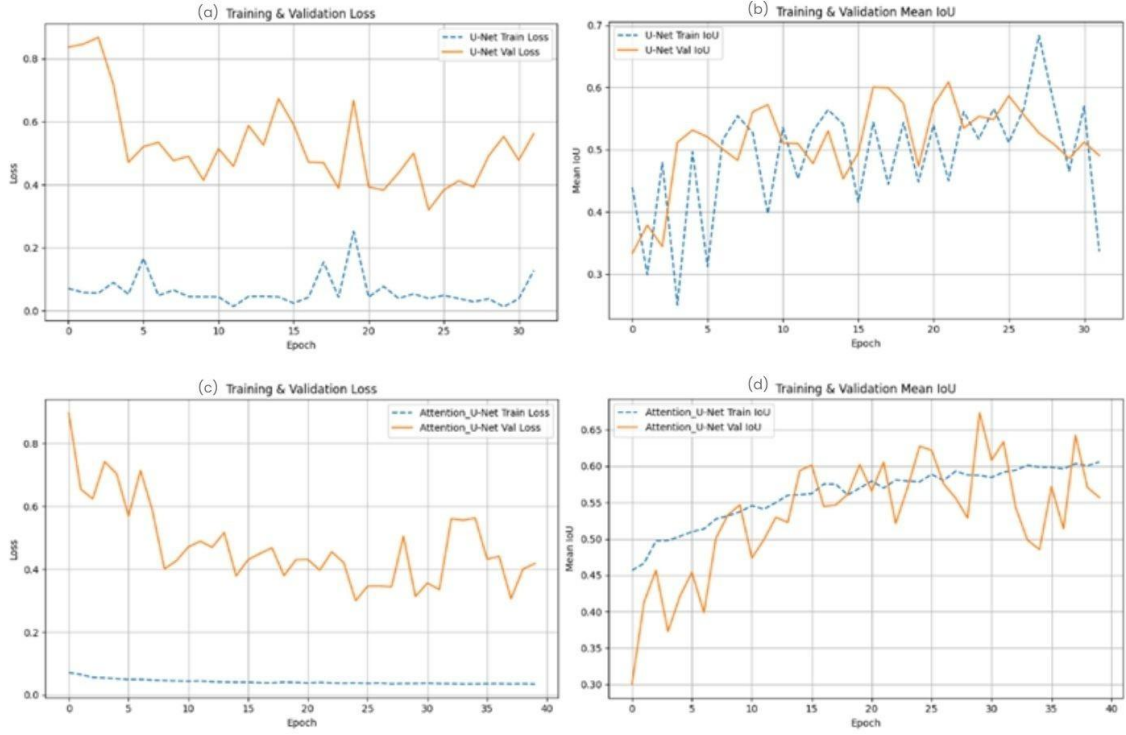
Figure 3. Comparison of training and validation learning curves for the standard U-Net and Attention U-Net models on the 4-class land cover task. (a) Training and validation loss for the U-Net. (b) Training and validation Mean IoU for the U-Net. (c) Training and

3.2 SEGMENTATION PERFORMANCE EVALUATION

Following training, the best saved models were evaluated on the held-out validation dataset. The primary goal of this evaluation was to assess the models' practical performance on the ultimate binary task of distinguishing "Forest" from "Non-Forest" areas. The models' 4-class predictions were reclassified into this binary scheme for the final analysis.

Table 1 provides a comparison of the binary classification metrics for both architectures. The Attention U-Net emerged as the superior model, achieving a higher overall pixel accuracy (93.47%) and a better-balanced F1-score for the "Forest" class (0.93). While the standard U-Net achieved an exceptionally high precision of 0.99 for the "Forest" class, this came at the cost of a significantly lower recall (0.85). The Attention U-Net demonstrated a more balanced and desirable performance, with a high precision of 0.96 and a substantially better recall of 0.90. For the goal of deforestation monitoring, where identifying all forest areas is critical, the Attention U-Net's higher recall makes it the more effective and reliable tool.

Table 1. Binary classification performance metrics for the U-Net and Attention U-Net models on the validation set. Metrics are reported for the "Forest" class.

Model	Precision	Recall	F1-Score	Overall Accuracy
U-Net	0.99	0.85	0.91	92.07%
Attention U-Net	0.96	0.90	0.93	93.47%

This trade-off is further clarified by the confusion matrices presented in Figure 4. The matrix for the standard U-Net shows a very low number of false positives for the 'Forest' class (Non-Forest pixels predicted as Forest), confirming its high precision. However, it also reveals a large number of false negatives (Forest pixels predicted as Non-Forest), explaining its lower recall. In contrast, the Attention U-Net's matrix shows a more balanced distribution of errors, reducing the number of false negatives at a slight cost to precision, ultimately resulting in a more useful model for the stated objective.
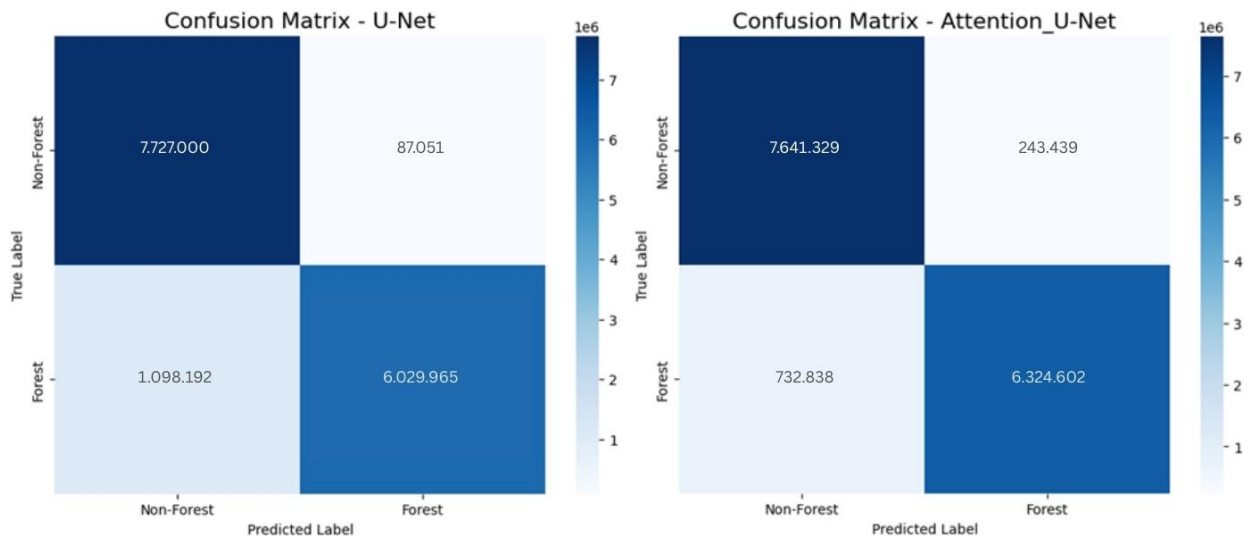

Figure 4. Confusion matrices for the binary (Forest vs. Non-Forest) classification task on the validation set for (a) the standard U-Net and (b) the Attention U-Net. Values represent the total number of pixels.

Visual inspection of the model predictions, as shown in Figure 5, further corroborates these quantitative findings. In Figure 5, the Attention U-Net (d) produces a visibly more accurate segmentation that aligns much more closely with the Ground Truth (b) than the standard U-Net (c). The most notable improvement is in the classification of transitional areas, where the standard U-Net tends to over-predict the 'Shrub/Scrub' class (yellow) along forest edges. The Attention U-Net, in contrast, provides a cleaner and more precise delineation between the

'Trees' and 'Grass' classes, resulting in a more coherent and realistic land cover map.
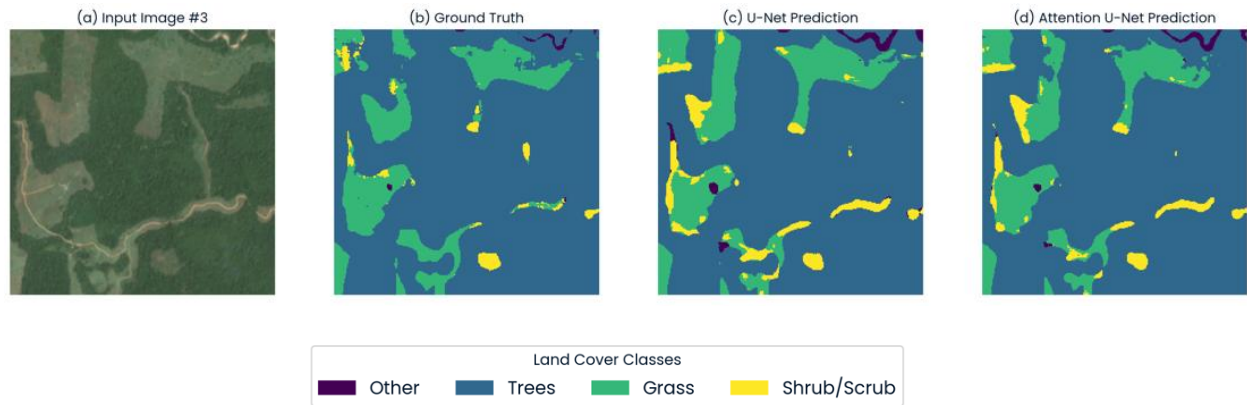

Figure 5. Qualitative comparison of segmentation results on sample patches from the validation set. (a) Input True-Color Image, (b) Ground Truth (Dynamic World), (c) U-Net Prediction, (d) Attention U-Net Prediction.

3.3 DEFORESTATION ANALYSIS IN THE SAN JOSÉ DEL GUAVIARE REGION

Given its superior performance, the trained Attention U-Net model was selected to conduct the final change detection analysis. The model was used to generate two land cover maps for the 4,430.23 km² Change Detection ROI, using cloud-free Sentinel-2 composites from the years 2021 (T1) and 2023 (T2). These 4-class maps were then reclassified into binary "Forest" and "Non-Forest" maps to identify areas of change.

The analysis revealed land cover transformation within the two-year period. As summarized in Table 2, the model detected a total of 232.15 km² of deforestation, representing a transition from "Forest" in 2021 to "Non-Forest" in 2023. Concurrently, 54.85 km² of reforestation or forest regrowth was identified. This results in a net forest loss of 177.30 km² within the study area. Figure 6 provides a visual example of the model's change detection capabilities on a single 256 x 256 pixel patch from the ROI. The 'Before' image (T1, 2021) on the left depicts a landscape dominated by dense forest cover, interspersed with several small, rectangular clearings characteristic of early-stage agricultural plots or pasture. The 'After' image (T2, 2023) in the centre shows a visible expansion of these clearings, with larger areas of exposed soil and low vegetation now apparent. The generated change map on the right translates these visual differences into a clear, quantitative output. The areas marked in red precisely correspond to the newly cleared plots, correctly identifying them as deforestation (Forest → Non-Forest). Furthermore, the model also identifies smaller, more subtle changes, such as the areas marked in blue, suggesting potential reforestation or natural regrowth in previously cleared areas. This granular, pixel-level analysis demonstrates the model's effectiveness in not only quantifying

large-scale change but also in capturing the complex, localized dynamics of the deforestation frontier.

Table 2. Quantified land cover change in the San José del Guaviare ROI between 2021 and 2023.

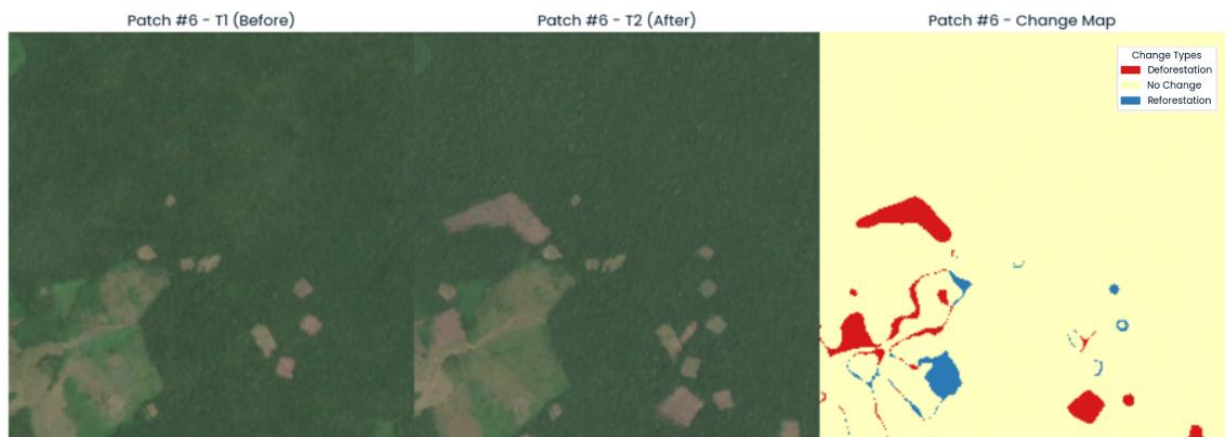| Change Type	Pixels Changed	Area (km²)
Deforestation	2,321,523	232.15
Reforestation	548,483	54.85
Net Change	-1,773,040	-177.30

Figure 6. An example of the change detection analysis on a sample patch. (Left) The landscape in 2021. (Center) The same landscape in 2023. (Right) The generated change map, where red indicates Deforestation (Forest -> Non-Forest), yellow indicates No Change, and blue indicates Reforestation (Non-Forest -> Forest).

4. DISCUSSION

This study successfully designed, implemented, and validated an end-to-end deep learning pipeline for monitoring deforestation in the Colombian Amazon. The results presented in the previous chapter not only demonstrate the technical viability of the approach but also offer significant insights into the performance of different neural network architectures and the dynamics of land cover change in a critical deforestation hotspot. This section interprets these findings, considers the methodological limitations, and outlines potential avenues for future work.

4.1 INTERPRETATION OF MODEL PERFORMANCE: THE EFFICACY OF ATTENTION MECHANISMS

The comparative analysis between the standard U-Net and the Attention U-Net yielded one of the study's principal findings: the integration of attention gates provided a distinct and measurable performance advantage. While both models were able to learn the 4-class segmentation task, the Attention U-Net achieved a higher peak validation Mean IoU (0.674 vs. 0.609) and demonstrated a more stable and prolonged learning progression (Figure 3). This suggests that the attention mechanism was highly effective at navigating the complex spectral signatures of the Colombian Amazon, enabling the model to learn more robust and generalizable features.

The final binary evaluation further illuminated the nature of this advantage. The standard U-Net, while achieving an exceptionally high precision of 0.99, did so at the cost of a low recall of 0.85 for the "Forest" class. As confirmed by its confusion matrix (Figure 4), this model adopted a strategy that aggressively minimized false positives. While this resulted in a very low rate of misclassifying "Non-Forest" land as "Forest," it came at the cost of a significant increase in false negatives, where true forest areas were frequently overlooked, particularly in fragmented or transitional zones. In contrast, the Attention U-Net struck a much more effective balance, achieving a high precision of 0.96 while substantially improving recall to 0.90. The visual evidence in Figure 5 corroborates this; the standard U-Net's tendency to over-predict "Shrub/Scrub" along forest edges is a classic sign of class confusion, which the Attention U-Net largely corrects. This indicates that the attention gates successfully guided the model to focus on the most salient features of the forest class, suppressing irrelevant information from adjacent land cover types and resulting in a more practically useful model for the primary goal of comprehensive forest monitoring.

4.2 SIGNIFICANCE OF THE CHANGE DETECTION ANALYSIS

The application of the trained Attention U-Net to the San José del Guaviare ROI successfully translated the model's predictive capabilities into actionable environmental intelligence. The quantification of a net forest loss of 177.30 km² between 2021 and 2023 provides a concrete, data-driven measure of the environmental pressure in this hotspot. This result offers granular, pixel-level insights that can be directly visualized and analysed (Figure 6).

The ability of this automated pipeline to produce such a quantitative analysis rapidly and over a large, inaccessible area demonstrates its significant practical value. Traditional monitoring methods would require extensive manual effort and would be unable to provide results with this level of detail and timeliness. This workflow, therefore, represents a scalable and efficient tool for conservation organizations and governmental bodies like IDEAM, enabling them to direct resources towards active deforestation fronts, verify the effectiveness of conservation policies, and better understand the spatial patterns of land cover change.

4.3 METHODOLOGICAL CONSIDERATIONS AND LIMITATIONS

Despite the promising results, it is essential to acknowledge the limitations inherent in the methodology. The most significant consideration is the reliance on the Dynamic World dataset as "pseudo ground truth." While it is a state-of-the-art global product and a highly effective proxy, it is itself the output of a model and is not free from error. It can sometimes struggle with highly heterogeneous landscapes or misclassify classes with similar spectral properties, and these inaccuracies are inevitably propagated to the model during training.

Secondly, the strategic decision to remap nine classes to four, while successful in overcoming the critical class imbalance issue, inherently involves a loss of ecological detail. Consolidating classes like "Built Area," "Crops," and "Bare Ground" into a single "Other" category was necessary for model stability but prevents a more detailed analysis of the specific drivers of deforestation (e.g., distinguishing forest-to-pasture conversion from forest-to-cropland).

Finally, the use of annual median composites, while a powerful technique for eliminating cloud cover, inherently masks intra-annual dynamics. Although Colombia does not experience distinct temperate seasons, its landscapes are subject to significant short-term changes driven by factors such as precipitation cycles, human activity, and agricultural calendars. For instance, a pasture might appear spectrally different during a drier period versus a wetter one, or a field used for short-cycle crops might be bare at one point in the year and vegetated at another. These nuances are averaged out in a yearly composite. Therefore, while this approach is highly

effective for detecting permanent land cover conversions like deforestation, it is less suited for monitoring more dynamic land use patterns or subtle phenological shifts within the year.

4.4 FUTURE DIRECTIONS

The success of this pipeline opens several exciting avenues for future research and development. To address the current limitations, the model could be enhanced by fusing the Dynamic World labels with other available ground-truth sources, such as regional data from local environmental agencies or manually verified points, to create a more robust and accurate training set.

Architecturally, the clear benefit of the attention mechanism suggests that further exploration of advanced models is warranted. An initial attempt was made to incorporate a pre-trained ResNet backbone via the DeepLabV3+ architecture; however, this resulted in immediate and catastrophic overfitting, as the powerful, pre-trained feature extractor quickly memorized the training data instead of learning generalizable patterns. This experience highlights that simply adopting a powerful backbone is insufficient; the architectural context in which it is placed is critical.

Therefore, a more promising future direction would be to test a hybrid U-Net that replaces its simple encoder with a pre-trained, efficient backbone like EfficientNet. This approach would synergize the proven feature extraction power of a modern classifier with the U-Net's superior decoder and skip-connection structure, which is well-suited for precise localization. Additionally, exploring architectures beyond the convolutional paradigm, such as the Transformer-based SegFormer, could yield further performance gains by better capturing long-range spatial context in the imagery.

From an application standpoint, the analysis could be extended to a finer temporal resolution, moving from annual to seasonal or quarterly composites. This would allow for a more dynamic monitoring system capable of providing earlier warnings of deforestation activities. Ultimately, the entire workflow could be operationalized and deployed on a cloud platform, creating a near-real-time monitoring tool that automatically ingests new Sentinel-2 imagery and flags areas of significant forest change, providing a powerful, data-driven asset in the ongoing effort to protect Colombia's vital forest ecosystems.

5. CONCLUSIONS

This study successfully developed and validated a scalable, automated deep learning pipeline for monitoring deforestation in the critically threatened Colombian Amazon. By leveraging the planetary-scale capabilities of Google Earth Engine and the globally consistent labels of the Dynamic World dataset, this work effectively addressed the persistent challenge of training data scarcity in remote sensing applications. The primary goals of this research, to design an automated data generation workflow, to compare state-of-the-art segmentation architectures, and to produce a quantitative analysis of forest change, were all successfully achieved.

The comparative analysis of neural network architectures yielded a clear and significant finding: the Attention U-Net demonstrated superior performance over the standard U-Net baseline. By integrating attention gates into its skip connections, the model learned to focus on the most salient features, resulting in more stable training, a higher 4-class validation Mean IoU (0.674 vs. 0.609), and a more effective balance between precision and recall on the final binary classification task. While the standard U-Net achieved high precision, its lower recall made it less suitable for the comprehensive monitoring required for conservation. The Attention U-Net, with a final Forest F1-score of 0.93 and a recall of 0.90, proved to be the more robust and reliable tool for the specific objective of identifying forest cover.

Ultimately, the application of the trained Attention U-Net model to a 4,430.23 km² Region of Interest (ROI) within the San José del Guaviare deforestation hotspot translated this technical success into actionable environmental insight. The analysis revealed a net forest loss of 177.30 km² within this specific ROI between 2021 and 2023, providing a granular, data-driven assessment of land cover change. This work concludes that the combination of a nuanced 4-class training strategy and an attention-based deep learning architecture provides a powerful and effective methodology for large-scale deforestation monitoring. The resulting pipeline represents a significant step forward, offering a scalable, cost-effective, and reproducible tool that can empower conservation efforts and inform policy interventions in the ongoing fight to protect Colombia's invaluable forest ecosystems.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The repository for this project is publicly available at https://github.com/esteban-russi/sentinel2-dl-deforestation.

REFERENCES

Alshehri, M., Ouadou, A. and Scott, G. J. (2024). 'Deep Transformer-Based Network Deforestation Detection in the Brazilian Amazon Using Sentinel-2 Imagery', IEEE Geoscience and Remote Remote Ssensing Lletters, 21. https://doi.org/10.1109/LGRS.2024.3355104.

Balaniuk, R., Isupova, O. and Reece, S. (2020). Mining and Tailings Dam Detection in Satellite Imagery Using Deep Learning. Sensors (Basel, Switzerland), 20(23), p.6936. https://doi.org/10.3390/s20236936.

Belgiu, M. and Drăguţ, L. (2016). 'Random forest in remote sensing: A review of applications and future directions', ISPRS Journal Of Photogrammetry And Remote Sensing, 114. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Brown, Christopher F, Steven P Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, et al. (2022). 'Dynamic World, Near real-time global 10 m land use land cover mapping', Scientific Data, 9 (1), p. 251. https://doi.org/10.1038/s41597-022-01307-4.

de Bem, Pablo, Osmar de Carvalho Junior, Renato Fontes Guimarães, and Roberto Trancoso Gomes. (2020). 'Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks.' Remote sensing (Basel, Switzerland), 12(6), p.901. https://doi.org/10.3390/rs12060901.

Drusch, M, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch. (2012). 'Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services', Remote Sensing of Environment, 120. https://doi.org/10.1016/j.rse.2011.11.026.

Ferreira, E, M Brito, R Balaniuk, M. S Alvim, and J. A dos Santos. (2020). 'Brazildam: A Benchmark Dataset For Tailings Dam Detection', International archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences., XLII-3/W12-2020 https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-343-2020.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore.. (2017). 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', Remote Sensing of Environment, 202. https://doi.org/10.1016/j.rse.2017.06.031.

Hansen, M. C, P. V Potapov, R Moore, M Hancher, S. A Turubanova, A Tyukavina, D Thau. (2013). 'High-Resolution Global Maps of 21st-Century Forest Cover Change', Science American Association for the Advancement of Science, 342 (6160). https://doi.org/10.1126/science.1244693.

IDEAM (2024). Monitoreo de la superficie de bosque y la deforestación en Colombia. Bogotá D.C: Instituto de Hidrología, Meteorología y Estudios Ambientales.

Isaienkov, Kostiantyn, Mykhailo Yushchuk, Vladyslav Khramtsov, and Oleg Seliverstov. (2021). Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2. IEEE journal of selected topics in applied earth observations and remote sensing, 14, pp.364–376. https://doi.org/10.1109/JSTARS.2020.3034186.

James, F. (2025). 'Remote Sensing for Deforestation Monitoring and Forest Management', Journal of Remote Sensing & GIS, J Remote Sens GIS.14:381. Vol.14 Iss.2 No:1000381

Kingma, D. P. and Ba, J. (2014). 'Adam: A Method for Stochastic Optimization'. https://doi.org/10.48550/arxiv.1412.6980.

Lecun, Y, L Bottou, Y Bengio, and P Haffner. (1998). 'Gradient-based learning applied to document recognition', Proceedings of the IEEE, 86 (11), pp. 2278-2324. https://doi.org/10.1109/5.726791.

Lv, Jinna, Qi Shen, Mingzheng Lv, Yiran Li, Lei Shi, and Peiying Zhang. (2023). 'Deep learning-based semantic segmentation of remote sensing images: a review', Frontiers in Ecology and Evolution, Volume 11 - 2023. https://doi.org/10.3389/fevo.2023.1201125.

Molina, M. (2024). 'Drivers of deforestation in the Colombian Amazon.' IUCN National Committee of The Netherlands. https://www.iucn.nl/app/uploads/2024/06/Drivers-of-deforestation-in-the-Colombian-Amazon_IUCN-NL-2024.pdf

Mountrakis, G., Im, J. and Ogole, C. (2011). 'Support vector machines in remote sensing: A review', ISPRS journal of photogrammetry and remote sensing, 66 (3). https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, et al. (2018). 'Attention U-Net: Learning Where to Look for the Pancreas'. https://doi.org/10.48550/arxiv.1804.03999.

Ortega Adarme, M. et al. (2022). Improving Deforestation Detection on Tropical Rainforests Using Sentinel-1 Data and Convolutional Neural Networks. Remote sensing (Basel, Switzerland), 14(14), p.3290. https://doi.org/10.3390/rs14143290.

Pal, M. (2005). 'Random forest classifier for remote sensing classification', International journal of remote sensing, 26 (1). https://doi.org/10.1080/01431160412331269698.

Panboonyuen, T., et al. (2021). 'Transformer-Based Decoder Designs for Semantic Segmentation on Remotely Sensed Images', Remote sensing (Basel, Switzerland), 13 (24). https://doi.org/10.3390/rs13245100.

Pekel, J.-F., et al. (2016). 'High-resolution mapping of global surface water and its long-term changes', Nature (London), 540 (7633), pp. 418-422. https://doi.org/10.1038/nature20584.

Procolombia (2023). 'Amazonia Region: Colombia's Green Lung'. Available at: https://colombia.co/en/colombia-country/geography-and-environment/amazonia-region (Accessed).

Rech, B. (2023). 'Effects of atmospheric correction on NDVI retrieved from Sentinel-2 imagery over different land cover classes', Anais do XX Simpósio Brasileiro de Sensoriamento Remoto.

Ronneberger, O., et al. (2015). 'U-Net: Convolutional Networks for Biomedical Image Segmentation', Lecture Notes in Computer Science. Switzerland: Springer International Publishing AG.

SIAT (2023). La Amazonia Colombiana: SIAT. Available at: https://siatac.co/la-amazonia-colombiana/ (Accessed: 7 July 2025 2025).

Torres, Daliana Lobo, Javier Noa Turnes, Pedro Juan Soto Vega, Raul Queiroz Feitosa, Daniel E Silva, Jose Marcato Junior, and Claudio Almeida. (2021). 'Deforestation Detection with Fully Convolutional Networks in the Amazon Forest from Landsat-8 and Sentinel-2 Images.' Remote sensing (Basel, Switzerland), 13(24), p.5084. https://doi.org/10.3390/rs13245084.

UNEP and FAO (2020). 'The State of the World's Forests 2020: Forests, Biodiversity and People'. Rome: Food and Agriculture Organization (FAO) and United Nations Environment Programme (UNEP). https://doi.org/https://doi.org/10.4060/ca8642en.

UNODC (2021). 'Colombia: Monitoreo de territorios afectados por cultivos ilícitos 2020.' Vienna, Austria.: United Nations Office on Drugs and Crime.

Wang, Z., et al. (2022). 'Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3', Computers & Geosciences, 158. https://doi.org/10.1016/j.cageo.2021.104969.

Wulder, Michael A, Joanne C White, Thomas R Loveland, Curtis E Woodcock, Alan S Belward, Warren B Cohen, Eugene A Fosnight, Jerad Shaw, Jeffrey G Masek, and David P Roy. (2016). 'The global

Landsat archive: Status, consolidation, and direction', Remote sensing of Environment, 185 pp. 271-283. https://doi.org/10.1016/j.rse.2015.11.032.

Zhu, Xiao Xiang, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. (2017). 'Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources', IEEE Geoscience and Remote Sensing Magazine, 5 (4). https://doi.org/10.1109/MGRS.2017.2762307.